

WRIGHT-FISHER-LIKE MODELS WITH CONSTANT POPULATION SIZE ON AVERAGE

NICOLAS GROSJEAN AND THIERRY HUILLET*

ABSTRACT. We first recall some basic facts from the theory of discrete-time Markov chains arising from two types neutral and non-neutral evolution models of population genetics with constant size. We then define and analyse a version of such models whose fluctuating total population size is conserved on average only. In our model, the population of interest is seen as being embedded in a frame process which is a critical Galton-Watson process. In this context, we address problems such as extinction, fixation, size of the population at fixation and survival probability to a bottleneck effect of the environment.

Running title: constant population size on average

Keywords: Markov chain population dynamics; Wright-Fisher-like models; constant population size on average; critical Galton-Watson process; extinction/fixation.

* corresponding author.

1. INTRODUCTION

Forward evolution of neutral large populations in genetics has a long history, starting in the 1920s; it is closely attached to the names of R. A. Fisher and S. Wright; see (T. Nagylaki, 1999) for historical commentaries. The book of (W. Ewens, 2004) is an excellent modern presentation of the current mathematical theory. The starting point of such neutral theories is embedded in the theory of discrete Markov chains whose transition matrices are obtained from branching Galton-Watson processes conditioned on keeping the total population size constant (as defined in Karlin-McGregor, 1964). Coalescent theory is the corresponding backward problem, obtained while running the forward neutral evolution processes backward-in-time. It was discovered independently by several researchers in the 1980s, but definitive formalization is commonly attributed to (J. Kingman, 1982). Major contributions to the development of coalescent theory were made (among others) by P. Donnelly, R. Griffiths, R. Hudson, F. Tajima and S. Tavaré (see the course of Tavaré in Saint-Flour 2004 for a review and references therein). The neutral theory has been enriched while including various drifts (or bias) describing say mutation, recombination, selection effects superposing to the genetic drift... Space-time scaling limits of such theories turn out to be very rich but in this manuscript we shall stick to the discrete space-time setting. All such recent developments and improvements concern chiefly the discrete neutral case and their various scaling limits in continuous time and/or space. As was shown for instance by (Möhle, 1994 and 1999), neutral forward and backward theories learn much from one another by using a concept

of duality introduced by (T. Liggett, 1985). There is therefore some evidence that the concept of duality could help one understand the backward theory even in non-neutral situations when various evolutionary forces are the causes of deviation to neutrality (J. Crow and M. Kimura, 1970); (T. Maruyama, 1977), (J. Gillespie, 1991) and (W. Ewens, 2004), for a discussion on various models of utmost interest in population genetics).

In this manuscript, after recalling some of the (forward and backward) theory of the discrete space-time neutral genetic drift with two types (or alleles), we briefly reconsider the case including various bias describing deviation to neutrality. Various examples are discussed to fix the background where we stress that the marginal populations of both types turn out to be Markovian, even in the biased case. As emphasized earlier, all this body of theory assumes an evolution process keeping constant over the generations the total size of the population. This condition that to form the next generation the offspring should preserve exactly the population size to constant is a drastic one and the purpose of this work is to discuss one way to understand a weaker form of conservation, namely conservation of the total population size on average only. And see how the previous theory is modified both in the neutral and the non-neutral cases with a bias included. In our approach, a critical Galton-Watson process plays a key role.

Let us summarize our results: as in the constant population size context, we first consider a two-types neutral population model in discrete-time. Type 1 population is seen as a subpopulation of a frame (or environment) process which is modeled as a critical Galton-Watson process whose size is constant over time, say n , but on average only. This process is already present in the Karlin-McGregor way to handle a constant population size n problem since, as n goes to infinity, such models boil down to a critical Galton-Watson process. The frame process exhibits very large growing fluctuations and it has long-range positive correlations. The type 1 population process also is a critical Galton-Watson process whose initial condition is a subset of size m of the n founders generating the frame process. We define the type 2 population as the one whose founders is made of the remaining subset of founders, therefore of size $n - m$. At all times therefore, the type 1 and 2 populations, as embedded sub-processes within the frame, sum up to the frame process itself. We then address the following problems in this setup: what is the extinction probability of type 1 population? And what is its fixation probability? For the former problem, a first point of view is the classical one for critical branching processes: extinction of type 1 population is when it first hits state 0. Such an extinction event occurs with probability 1 but it takes a very long time to do so. For the latter problem, we say that a type 1 fixation event occurs at some time if this time is the first at which type 2 population goes extinct while type 1 population remains alive. We compute the probability distribution of the fixation time, together with the probability that a fixation event takes place. We find that the fixation time has finite mean (of order n when both m, n are large or when m is fixed and n large), whereas its variance is infinite. The fixation probability itself is found to be of order m/n under the same assumptions.

Then we address the following problem: what is the size of type 1 population at fixation whenever this event occurs? We find that on average it is m , while its variance is infinite. Finally we discuss an alternative and more symmetric way of

defining type 1 population extinction as the fixation of the type 2 population in the latter sense. With these definitions at hand, one can compute the probability that a fixation event of type 1 precedes an extinction event and conversely.

We then observe that, under the neutral hypothesis, type 1 and frame processes, although not independent, are both Markov processes both marginally and jointly. Inspired by similar ideas in the context of constant population size models, we then introduce bias (as a deviation to neutrality) into our model with fluctuating total population size. We show that with this bias included, type 1 and frame processes (still the same critical mean- n Galton-Watson process as in the neutral case) are still jointly Markov but the type 1 marginal process no longer is Markovian.

Finally, we briefly discuss the problem of defining an effective population size in our model with variable population size and we investigate the related question of computing the type 1 survival probability to a bottleneck effect of the frame process. In this context, the critical homographic Galton-Watson model, which is invariant under iterated composition, is shown to be of particular interest.

2. DISCRETE-TIME NEUTRAL GENETIC DRIFT AND COALESCENT: A REMINDER

In this Section, to fix the background and notations, we review some well-known facts from the cited literature.

2.1. Exchangeable neutral population models: Reproduction laws examples. (The Cannings, 1974 model). Consider a population with non-overlapping generations $r \in \mathbb{Z}$. Assume the population size is constant, say with n individuals (or genes) over the generations. Assume the random reproduction law at generation 0 is $\nu_n := (\nu_{1,n}, \dots, \nu_{n,n})$, satisfying:

$$\sum_{m=1}^n \nu_{m,n} = n.$$

Here, $\nu_{m,n}$ is the number of offspring of gene m . We avoid the trivial case: $\nu_{m,n} = 1$, $m = 1, \dots, n$. One iterates the reproduction over generations, while imposing the following additional assumptions:

- Exchangeability: $(\nu_{1,n}, \dots, \nu_{n,n}) \stackrel{d}{=} (\nu_{\sigma(1),n}, \dots, \nu_{\sigma(n),n})$, for all n -permutations $\sigma \in \mathcal{S}_n$.

- time-homogeneity: reproduction laws are independent and identically distributed (iid) at each generation $r \in \mathbb{Z}$.

This model therefore consists of a conservative conditioned branching Galton-Watson process in $[n]^{\mathbb{Z}}$, where $[n] := \{0, 1, \dots, n\}$ (see Karlin-McGregor, 1964).

Famous reproduction laws are:

Example 1 The θ -multinomial Dirichlet family: $\nu_n \stackrel{d}{\sim} \text{Multin-Dirichlet}(n; \theta)$, where $\theta > 0$ is some ‘disorder’ parameter. With $\mathbf{k}_n := (k_1, \dots, k_n)$, ν_n admits the following joint exchangeable distribution on the simplex $|\mathbf{k}_n| := \sum_{m=1}^n k_m = n$:

$$\mathbf{P}(\nu_n = \mathbf{k}_n) = \frac{n!}{[n\theta]_n} \prod_{m=1}^n \frac{[\theta]_{k_m}}{k_m!},$$

where $[\theta]_k = \theta(\theta+1)\dots(\theta+k-1)$ is the rising factorial of θ . This distribution can be obtained by conditioning n independent mean 1 Pòlya distributed random variables $\boldsymbol{\xi}_n = (\xi_1, \dots, \xi_n)$ on summing to n , that is to say: $\boldsymbol{\nu}_n \stackrel{d}{=} (\boldsymbol{\xi}_n : |\boldsymbol{\xi}_n| = n)$, where, with $\mathbb{N}_0 = \{0, 1, \dots\}$,

$$\mathbf{P}(\xi_1 = k) = \frac{[\theta]_k}{k!} (1+\theta)^{-k} (\theta/(1+\theta))^\theta, \quad k \in \mathbb{N}_0$$

or, equivalently in terms of its probability generating function (pgf),

$$\mathbf{E}(z^{\xi_1}) =: \phi(z) = \left(1 - \frac{1}{\theta}(z-1)\right)^{-\theta}, \quad z < z_c := 1 + \theta.$$

Pòlya (or negative binomial) distributed random variables are known to be compound-Poisson (or infinitely divisible), meaning (Steutel and van Harn, 2003)

$$\phi(z) = e^{-\lambda(1-\psi(z))},$$

with $\lambda = \theta \log(1 + 1/\theta) > 0$ and with $\psi(z)$ a pgf obeying $\psi(0) = 0$ and $\psi'(1) = 1/\lambda$.

When $\theta = 1$, ξ_1 is geometric and $\boldsymbol{\nu}_n$ is uniformly distributed on the simplex $\mathbf{k}_n : |\mathbf{k}_n| = n$, with $\mathbf{P}(\boldsymbol{\nu}_n = \mathbf{k}_n) = 1/\binom{2n-1}{n}$.

When $\theta \rightarrow \infty$, this distribution reduces to the Wright-Fisher model for which $\boldsymbol{\nu}_n \stackrel{d}{\sim} \text{Multin}(n; 1/n, \dots, 1/n)$. Indeed, $\boldsymbol{\nu}_n$ admits the following joint exchangeable multinomial distribution on the simplex $\mathbf{k}_n : |\mathbf{k}_n| = n$:

$$\mathbf{P}(\boldsymbol{\nu}_n = \mathbf{k}_n) = \frac{n! \cdot n^{-n}}{\prod_{m=1}^n k_m!}.$$

This distribution can be obtained by conditioning n independent mean 1 Poisson distributed random variables $\boldsymbol{\xi}_n = (\xi_1, \dots, \xi_n)$ on summing to n . When n is large, using Stirling formula, it follows that $\boldsymbol{\nu}_n \xrightarrow[n \rightarrow \infty]{d} \boldsymbol{\xi}_\infty$ with joint finite-dimensional Poisson law: $\mathbf{P}(\boldsymbol{\xi}_n = \mathbf{k}_n) = \prod_{m=1}^n \frac{e^{-1}}{k_m!} = \frac{e^{-n}}{\prod_{m=1}^n k_m!}$ on \mathbb{N}_0^n . Thanks to the product form of all finite-dimensional laws of $\boldsymbol{\xi}_\infty$, we get an asymptotic independence property of $\boldsymbol{\nu}_n$.

A slight extension of the Pòlya model would be to consider a compound Pòlya model (and compound Poisson as well) for which, for some $z_c > 1$,

$$\mathbf{E}(z^{\xi_1}) =: \phi(z) = \left(1 - \frac{1}{\theta h'(1)}(h(z) - 1)\right)^{-\theta}, \quad z < z_c,$$

with $h(z)$ some pgf obeying $h(0) = 0$, $h'(1) > 1$ and $h''(1) < \infty$. Here, as $\theta \rightarrow \infty$, ξ_1 is compound Poisson with rate $1/h'(1)$.

Example 2 Take for ξ the model: $\xi = 0$ with probability $1/2$, $\xi = 2$ with probability $1/2$, so with $\phi(z) = (1 + z^2)/2$. Here, provided n is even,

$$\mathbf{P}(\boldsymbol{\nu}_n = \mathbf{k}_n) = 1/\binom{n}{n/2}, \quad k_m \in \{0, 2\}, \quad m = 1, \dots, n, \text{ uniform on } |\mathbf{k}_n| = n.$$

Example 3 In the Moran model, $\boldsymbol{\nu}_n \stackrel{d}{\sim}$ random permutation of $(2, 0, 1, \dots, 1)$: in such a model, only one new gene per generation may come to life, at the expense

of the simultaneous disappearance of some other gene.

In the first two examples, ν_n is obtained while conditioning n iid mean 1 random variables ξ_n on summing to n (not the case of Moran model) and we shall only consider this case in the sequel, even assuming (as in the examples) the ξ_m s to have finite variance at least: $\sigma^2(\xi) = \phi''(1) < \infty$. In all such cases,

$$\begin{aligned} \mathbf{P}(\nu_n = \mathbf{k}_n) &= \frac{\prod_{m=1}^n [z^{k_m}] \phi(z)}{[z^n] \phi(z)^n}, \quad |\mathbf{k}_n| = n \\ \mathbf{P}(\nu_{1,n} + \dots + \nu_{k,n} = k') &= \frac{[z^{k'}] \phi(z)^k [z^{n-k'}] \phi(z)^{n-k}}{[z^n] \phi(z)^n}, \quad k, k' \in \{0, \dots, n\} \\ \mathbf{E}(z^{\nu_{1,n}}) &= \frac{1}{[z^n] \phi(z)^n} \sum_{k'=0}^n z^{k'} [z^{k'}] \phi(z) [z^{n-k'}] \phi(z)^{n-1}. \end{aligned}$$

2.2. Forward in time branching process (neutral genetic drift). Take a sub-sample of size m out of $[n] := \{0, 1, \dots, n\}$, at generation 0. Given $N_0(m) = m$, let

$N_r(m) = \#$ offspring at generation $r \in \mathbb{N}_0$, forward-in-time.

$N_r(m)$ represents the descendance at time r of the (say type 1) m first individuals of the whole population of size n . This sibship process is a discrete-time homogeneous Markov chain, with transition probability:

$$(1) \quad \mathbf{P}(N_{r+1}(m) = k' \mid N_r(m) = k) = \mathbf{P}(\nu_{1,n} + \dots + \nu_{k,n} = k').$$

It is a martingale, with state-space $\{0, \dots, n\}$, initial state m , absorbing states $\{0, n\}$ and transient states $\{1, \dots, n-1\}$. With $\tau_{m,0} = \inf(r : N_r(m) = 0)$ and $\tau_{m,n} = \inf(r : N_r(m) = n)$, the first hitting time of the boundaries $\{0, n\}$ is: $\tau_m = \tau_{m,0} \wedge \tau_{m,n}$. It is finite with probability 1 and has finite mean. Omitting reference to any specific initial condition m , the process $(N_r; r \in \mathbb{N}_0)$ has the transition matrix Π_n with entries $\Pi_n(k, k') = \mathbf{P}(\nu_{1,n} + \dots + \nu_{k,n} = k')$ given by (1). We have $\Pi_n(0, k') = \delta_{0,k'}$ and $\Pi_n(n, k') = \delta_{n,k'}$ and Π_n is not irreducible. However, Π_n is aperiodic and (apart from absorbing states) cannot be broken down into non-communicating subsets; as a result it is diagonalizable, with eigenvalues $\lambda_0 \geq \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ and $1 = \lambda_0 = \lambda_1 > \lambda_2$. For $m, k \in \{0, n\}$, we have

$$\mathbf{P}(N_r(m) = k) = \mathbf{e}_m' \Pi_n^r \mathbf{e}_k$$

and therefore, with $\bar{\Pi}_n$ the restriction of Π_n to the states $\{1, n-1\}$, for $m \in \{1, n-1\}$

$$\mathbf{P}(\tau_m > r) = \mathbf{e}_m' \bar{\Pi}_n^r \mathbf{1},$$

where $\mathbf{1}$ is the all-one column vector. τ_m has geometric tails with rate $\lambda_2 < 1$.

Example 4 (Dirichlet binomial): With U_k a $(0, 1)$ -valued random variable with density $\text{beta}(k\theta, (n-k)\theta)$

$$\mathbf{P}(\nu_{1,n} + \dots + \nu_{k,n} = k') = \binom{n}{k'} \frac{[k\theta]_{k'} [(n-k)\theta]_{n-k'}}{[n\theta]_n} = \mathbf{E} \left[\binom{n}{k'} U_k^{k'} (1 - U_k)^{n-k'} \right],$$

which is a beta mixture of the binomial distribution $\text{Bin}(n, u)$. In particular, with U_1 a $(0, 1)$ -valued random variable with density $\text{beta}(\theta, (n-1)\theta)$, we have

$$(2) \quad \mathbf{E}(z^{\nu_{1,n}}) = \mathbf{E}[(zU_1 + 1 - U_1)^n].$$

Example 5 The Wright-Fisher model has a $\text{Bin}(n, k/n)$ transition matrix:

$$\mathbf{P}(N_{r+1}(m) = k' \mid N_r(m) = k) = \binom{n}{k'} \left(\frac{k}{n}\right)^{k'} \left(1 - \frac{k}{n}\right)^{n-k'}.$$

Remarks 1-3

1/- (statistical symmetry): Due to exchangeability of the reproduction law, neutral models are symmetric in the following sense: The transition probabilities of $\overline{N}_r(m) := n - N_r(m)$ are equal to the transition probabilities of $N_r(m)$. $\overline{N}_r(m) \stackrel{d}{=} N_r(n-m)$ represents the descendance at time r of the (say type 2) $n-m$ initial individuals complementing m to get the whole initial population of size n .

2/- The smaller n , the more the genetic drift process $N_r(m)$ looks chaotic as it will reach very fast one of the absorbing boundary (genetic diversity is reduced fast); the smaller the population the greater the probability that fluctuations will lead to absorption. For varying population sizes n , to define an effective (equivalent) population size quantifying the genetic drift, more weight therefore has to be put on the small values of n than on the large ones where the process evolves more smoothly.

3/- As $n \rightarrow \infty$, the finite-dimensional probability transition matrix $\Pi_n(k, k') = \mathbf{P}(\nu_{1,n} + \dots + \nu_{k,n} = k')$ (with $k, k' \in [n]$), given by (1), tends to the infinite-dimensional transition matrix Π with (k, k') entries, $k, k' \in \mathbb{N}_0$

$$\mathbf{P}(N_{r+1}(m) = k' \mid N_r(m) = k) = \Pi(k, k') = \mathbf{P}(\xi_1 + \dots + \xi_k = k') = [z^{k'}] \phi(z)^k.$$

The latter is the one of a critical branching Galton-Watson process with offspring distribution ξ given by its pgf $\phi(z) = \mathbf{E}(z^\xi)$. \square

2.3. Backward in time process (neutral coalescent). The coalescent backward count process can be defined as follows: Take a sub-sample of size m from $[n]$ at generation 0. Identify two individuals from $[m]$ at each step if they share a common ancestor one generation backward-in-time. This defines an equivalence relation between 2 genes from the set $[m]$. With $A_0(m) = m$, let

$$A_r(m) = \# \text{ ancestors at generation } r \in \mathbb{N}_0, \text{ backward-in-time.}$$

The backward ancestral count process is a discrete-time Markov chain with transition probabilities (Cannings, 1974 and Gladstien, 1978):

$$(3) \quad \mathbf{P}(A_{r+1}(m) = a \mid A_r(m) = b) = P_{b,a}^{(n)} := \frac{b!}{a!} \sum_{b_1, \dots, b_a}^* \frac{P_{b;a}^{(n)}(\mathbf{b}_a)}{b_1! \dots b_a!}.$$

$$= \frac{\binom{n}{a}}{\binom{n}{b}} \sum_{b_1, \dots, b_a}^* \mathbf{E} \left(\prod_{l=1}^a \binom{\nu_{l,n}}{b_l} \right).$$

In the latter equations, the star-sum is over $b_1, \dots, b_a \in \mathbb{N} := \{1, 2, \dots\}$, such that $b_1 + \dots + b_a = b$. This Markov chain has state-space $\{0, \dots, m\}$, initial state m , absorbing states $\{0, 1\}$. The process $(A_r; r \in \mathbb{N}_0)$ has the transition matrix P_n with entries $P_n(b, a) = P_{b,a}^{(n)}$ given by (3). In particular, the probability that two (three) randomly sampled without replacement individuals out of $A_r(m)$ share a common ancestor in the previous generation are, respectively,

$$P_{2,1}^{(n)} = \frac{n \mathbf{E} \left(\binom{\nu_{1,n}}{2} \right)}{\binom{n}{2}}, \quad P_{3,1}^{(n)} = \frac{n \mathbf{E} \left(\binom{\nu_{1,n}}{3} \right)}{\binom{n}{3}}.$$

If $P_{3,1}^{(n)}/P_{2,1}^{(n)} \rightarrow 0$ as $n \rightarrow \infty$ and $P_{2,1}^{(n)} = O(n^{-1})$, the equivalent population size is $n_e = 1/P_{2,1}^{(n)} = O(n)$. It can be shown (see e.g. Huillet and Möhle, Theorem 2.4, 2015) that this situation occurs under the condition $\sigma^2(\xi) = \phi''(1) < \infty$, and then $n_e \sim n/\sigma^2(\xi)$ for large n . In the Pòlya case for example, $P_{3,1}^{(n)}/P_{2,1}^{(n)} \rightarrow 0$ as $n \rightarrow \infty$ and using (2), $P_{2,1}^{(n)} = (\theta + 1)/(n\theta + 1)$ leading to $n_e = (n\theta + 1)/(\theta + 1)$ with $n_e \sim n/\sigma^2(\xi)$ and $\sigma^2(\xi) = (\theta + 1)/\theta$. Under such conditions, as is well-known, a space-time scaling limit exists for $N_r(m)$ (the Wright-Fisher diffusion on the unit interval), together with a time-scaled version of $A_r(m)$ (the Kingman coalescent), with continuous time measured in units of n_e . So n_e in the context of the neutral forward dynamics is defined from the backward setup and it fixes the true time-scale.

2.4. Duality (neutral case). We start with a definition of the duality concept which is relevant to our purposes.

Definition (Liggett, 1985): Two Markov processes $(X_t^1, X_t^2; t \geq 0)$, with state-spaces $(\mathcal{E}_1, \mathcal{E}_2)$, are said to be dual with respect to some real-valued function Φ on the product space $\mathcal{E}_1 \times \mathcal{E}_2$ if $\forall x_1 \in \mathcal{E}_1, \forall x_2 \in \mathcal{E}_2, \forall t \geq 0$:

$$(4) \quad \mathbf{E}_{x_1} \Phi(X_t^1, x_2) = \mathbf{E}_{x_2} \Phi(x_1, X_t^2).$$

We then recall basic examples of dual processes from the neutral and exchangeable population models (Möhle, 1997): The neutral forward and backward processes $(N_r, A_r; r \in \mathbb{N}_0)$ introduced in the two preceding subsections are dual with respect to the hypergeometric sampling without replacement kernel:

$$(5) \quad \Phi_n(m, k) = \binom{n-m}{k} / \binom{n}{k} \text{ on } \{0, \dots, n\}^2.$$

(5) reads:

$$\mathbf{E}_m \left[\binom{n-N_r}{k} / \binom{n}{k} \right] = \mathbf{E}_k \left[\binom{n-m}{A_r} / \binom{n}{A_r} \right] = \mathbf{E}_k \left[\binom{n-A_r}{m} / \binom{n}{m} \right].$$

The left-hand-side is the probability that a k -sample (without replacement) from population of size N_r at time r are all of type 2, given $N_0 = m$. If this k -sample are all descendants of A_r ancestors at time $-r$, this probability must be equal to the probability that a m -sample from population of size A_r at time $-r$ are themselves all of type 2.

With P'_n the transpose of P_n , a one-step ($r = 1$) version of these formulae is:

$$\Pi_n \Phi_n = \Phi_n P'_n$$

where Φ_n is an $(n+1) \times (n+1)$ matrix with entries $\Phi_n(m, k)$ and (Π_n, P_n) the transition matrices of forward and backward processes. Note that the matrix Φ_n is symmetric and left-upper triangular. The matrix Φ_n is invertible, with entries

$$\Phi_n^{-1}(i, j) = (-1)^{i+j-n} \binom{i}{n-j} \binom{n}{i} = (-1)^{i+j-n} \binom{j}{n-i} \binom{n}{j}.$$

The matrix Φ_n^{-1} is symmetric right-lower triangular. Thus,

$$\Phi_n^{-1} \Pi_n \Phi_n = P'_n.$$

Being similar matrices, Π_n and P'_n (or P_n) both share the same eigenvalues. In (Möhle, 1999), a direct combinatorial proof of the duality result can be found (in the general exchangeable or neutral case); it was obtained by directly checking the consistency of (1), (3) and (5).

The duality formulae allow one to deduce the probabilistic structure of one process from the one of the other.

3. BEYOND NEUTRALITY (SYMMETRY BREAKING ARISING FROM BIAS)

Discrete forward non-neutral Markov chain models (with non-null drifts) can be obtained by substituting

$$k \rightarrow np \left(\frac{k}{n} \right) \text{ in } \Pi_n(k, k') := \mathbf{P}(\nu_{1,n} + \dots + \nu_{k,n} = k'),$$

where:

$$p(x) : x \in (0, 1) \rightarrow (0, 1) \text{ is continuous, increasing, with } p(0) = 0, p(1) = 1.$$

$p(x)$ is the state-dependent Bernoulli bias probability different from the identity x (as in neutral case).

When particularized to the Wright-Fisher model, this leads to the biased transition probabilities:

$$\mathbf{P}(N_{r+1}(m) = k' \mid N_r(m) = k) = \binom{n}{k'} p \left(\frac{k}{n} \right)^{k'} \left(1 - p \left(\frac{k}{n} \right) \right)^{n-k'}.$$

In this binomial n -sampling with replacement model, a type 1 individual is drawn with probability $p \left(\frac{k}{n} \right)$ which is different from the uniform distribution k/n , due to bias effects.

From this, we conclude (a symmetry breaking property): The transition probabilities of $\overline{N}_r(m) := n - N_r(m)$, $r \in \mathbb{N}_0$ are

$$\text{Bin}(n, 1 - p(1 - k/n)) \neq \text{Bin}(n, p(k/n)),$$

and so, no longer coincide with the ones of $(N_r(m); r \in \mathbb{N}_0)$. The process $N_r(m)$, $r \in \mathbb{N}_0$, which is Markovian, no longer is a martingale. Rather, if $x \rightarrow p(x)$ is concave (convex), $N_r(m)$, $r \in \mathbb{N}_0$ is a sub-martingale (super-martingale), because: $\mathbf{E}(N_{r+1}(m) \mid N_r(m)) = np(N_r(m)/n) \geq N_r(m)$ (respectively $\leq N_r(m)$).

In the binomial neutral Wright-Fisher transition probabilities for instance, we replaced the success probability k/n by a more general function $p(k/n)$. The reproduction law corresponding to the biased binomial model is multinomial and asymmetric, namely: $\nu_n \stackrel{d}{\sim} \text{Multin}(n; \pi_n)$, where $\pi_n := (\pi_{1,n}, \dots, \pi_{n,n})$ and: $\pi_{m,n} = p(m/n) - p((m-1)/n)$, $m = 1, \dots, n$ obeying $\sum_{m=1}^n \pi_{m,n} = p(1) - p(0) = 1$. Due to its asymmetry, the law of the biased ν_n no longer is exchangeable. We now recall some well-known bias examples arising in population genetics.

Example 6 (homographic model, haploid selection). Assume

$$(6) \quad p(x) = (1+s)x / (1+sx),$$

where $s > -1$ is a selection parameter. This model arises when gene 1 (respectively 2), with frequency x (respectively $1-x$), has fitness $1+s$ (respectively 1) in a multiplicative model of fitness. The case $s > 0$ arises when gene of type 1 is selectively advantageous, whereas it is disadvantageous when $s \in (-1, 0)$.

Example 7 (diploid selection with dominance). Assume

$$(7) \quad p(x) = \frac{(1+s)x^2 + (1+sh)x(1-x)}{1+sx^2 + 2shx(1-x)}.$$

In this model, genotype 11 (respectively 12 and 22), with frequency x^2 (respectively $2x(1-x)$ and $(1-x)^2$) has fitness $1+s$ (respectively $1+sh$ and 1). h is a measure of the degree of dominance of heterozygote 12. We impose $s > -1$ and $sh > -1$. Note that the latter quantity can be put into the canonical form of deviation to neutrality:

$$p(x) = x + sx(1-x) \frac{h - x(2h-1)}{1+sx^2 + 2shx(1-x)},$$

where the ratio appearing in the right-hand-side is the ratio of the difference of marginal fitnesses of 1 and 2 to their mean fitness. The case $h = 1/2$ corresponds to balancing selection with: $p(x) = x + \frac{s}{2} \frac{x(1-x)}{1+sx}$.

Example 8 (quadratic segregation model) With $a \in [-1, 1]$, a curvature parameter, one may choose:

$$(8) \quad p(x) = x(1+a-ax),$$

corresponding to a segregation model (Weissing and van Boven, 2001). If $a = 1$, $p(x) = x(2-x) = 1 - (1-x)^2$: this bias appears in a discrete 2-sex population model (Möhle, 1994).

We can relax the assumption $p(0) = 0, p(1) = 1$ by assuming $0 \leq p(0) \leq p(1) \leq 1$, $p(1) - p(0) \in [0, 1]$.

Example 9 (affine mutation model) Take for example

$$(9) \quad p(x) = (1-\mu_2)x + \mu_1(1-x),$$

where (μ_1, μ_2) are mutation probabilities, satisfying $\mu_1 \leq 1 - \mu_2$. It corresponds to the mutation scheme: $1 \xrightleftharpoons[\mu_2]{\mu_1} 2$. To avoid discussions of intermediate cases, we will assume that $p(0) = \mu_1 > 0$ and $p(1) < 1$ ($\mu_2 > 0$). In this case, the matrix Π_n is irreducible and even primitive and all states of this Markov chain are now recurrent. We have $\mathbf{P}(N_{r+1} > 0 \mid N_r = 0) = 1 - (1 - p(0))^n > 0$ and

$\mathbf{P}(N_{r+1} < n \mid N_r = n) = 1 - p(1)^n > 0$ and the boundaries $\{0\}$ and $\{n\}$ no longer are strictly absorbing as there is a positive reflection probability inside the domain $\{0, 1, \dots, n\}$.

4. A MODIFIED CANNINGS MODEL WITH CONSTANT POPULATION SIZE ON AVERAGE

The condition that to form the next generation the offspring should preserve exactly the population size to n is a drastic one and we wish here to discuss a weaker form of conservation, namely conservation of the total population size on average. And see how the latter theory with constant population size is modified in depth.

The frame (or environment) process. Consider a population of initial size n as before. Assume each individual generates a random number of offspring of size ξ_m , independently of one another, with the ξ_m s iid with $\mathbf{P}(\xi = k) =: \pi_k$. Assume $\mathbf{E}(\xi) = 1$ and introduce $\mathbf{E}(z^\xi) =: \phi(z)$, the pgf of ξ then with $\phi'(1) = 1$. We shall moreover assume $\sigma^2(\xi) = \phi''(1) < \infty$ (finiteness of the variance of ξ). Then we are left with a classical critical branching Galton-Watson process with finite variance.

Let $N_r(n)$ be the number of descendants at generation r , given initially $N_0(n) = n$. We have

$$(10) \quad N_{r+1}(n) = \sum_{m=1}^{N_r(n)} \xi_m,$$

so that $\mathbf{E}(N_{r+1}(n)) = \mathbf{E}(\xi) \mathbf{E}(N_r(n))$. Thus for each $r \geq 0$, $\mathbf{E}(N_r(n)) = n$ and the population size is conserved on average only. We shall call the process $N_r(n)$, $r \geq 0$, the frame or environment process. It models a fluctuating total population size constant on average. Note that $n = N_0$ could be made random with mean μ so that in this case $\mathbf{E}(N_r(N_0)) = \mathbf{E}(N_0) = \mu$ but in the sequel, for the sake of simplicity, we shall work conditionally given $N_0 = n$.

A two types population model. Let $1 \leq m \leq n$ and $N_r(m)$ be the number of (say type 1) descendants at generation r of the $N_0(m) = m$ first founders (as a subset of the full initial population with n founders). We shall also let $N_r(n - m) = N_r(n) - N_r(m)$ be the number of (say type 2) descendants at generation r of the remaining part of the initial population. With $\phi_r(z) := \mathbf{E}(z^{N_r(1)})$ the pgf of $N_r(1)$, we have $\phi_{r+1}(z) = \phi(\phi_r(z))$, $\phi_0(z) = z$ so that $\phi_r(z) = \phi^{\circ r}(z)$ the r -fold composition of ϕ with itself. And by independence of the founders

$$(11) \quad \begin{cases} \mathbf{E}(z^{N_r(m)}) = \phi_r(z)^m, \text{ with} \\ \mathbf{E}(N_r(m)) = m \text{ and } \mathbf{E}(N_r(m)^2) = m^2 + rm\phi''(1). \end{cases}$$

If $r_2 > r_1 > 0$, using the following expression of the joint pgf of $N_{r_1}(m)$, $N_{r_2}(m)$, say $\mathbf{E}(z_1^{N_{r_1}(m)} z_2^{N_{r_2}(m)}) = \phi_{r_1}(z_1 \phi_{r_2-r_1}(z_2))^m$, we get

$$\begin{aligned} \text{Cov}(N_{r_1}(m), N_{r_2}(m)) &= r_1 m \phi''(1), \text{ independently of } r_2 > r_1 \\ \text{Corr}(N_{r_1}(m), N_{r_2}(m)) &= \frac{\text{Cov}(N_{r_1}(m), N_{r_2}(m))}{\sigma(N_{r_1}(m)) \sigma(N_{r_1}(m))} = \sqrt{\frac{r_1}{r_2}} = \left(1 + \frac{r_2 - r_1}{r_1}\right)^{-1/2}, \end{aligned}$$

independent of m . The process $N_r(m)$ exhibits positive long-range correlations.

Note that if $m = n$, we are left with the frame process and $m < n \Rightarrow N_r(m) \leq N_r(n)$, $\forall r$: $N_r(m)$ is a sub-process of $N_r(n)$. While if $m = 1$, the process $N_r(1)$, as a special sub-process of $N_r(n)$, describes the fate of an initial mutant.

The process $N_r(m)$, $r \in \mathbb{N}_0$, as a critical Galton-Watson branching process itself, is a discrete-time Markov chain with initial condition $N_0(m) = m$, state-space \mathbb{N}_0 , and (semi-)infinite-dimensional transition matrix

$$\mathbf{P}(N_{r+1}(m) = k' \mid N_r(m) = k) = \left[z^{k'} \right] \phi(z)^k =: \Pi, \quad k, k' \in \mathbb{N}_0.$$

Therefore, with $\mathbf{e}'_m = (0, \dots, 0, 1, 0, \dots, 0)$ with the 1 in position m , $m = 0, 1, \dots$ ($'$ denoting transposition of the column vector \mathbf{e}_m),

$$\mathbf{P}(N_r(m) = k) = \mathbf{e}'_m \Pi^r \mathbf{e}_k.$$

The process $N_r(m)$ is the process of interest to us, as a process embedded in the frame process $N_r(n)$ giving the total population size which is highly fluctuating, although with constant mean n .

Type 1 extinction event. The type 1 descendants of the m original founders go extinct with probability 1 because this probability is ρ_r^m where $\rho_r := \mathbf{P}(N_r(1) = 0) = \phi_r(0)$ tends (slowly) to 1 as $r \rightarrow \infty$ (for a critical Galton-Watson process, extinction is almost sure and state 0 is absorbing for $N_r(m)$). Indeed, ρ_r obeys

$$\rho_{r+1} = \phi(\rho_r), \quad \rho_0 = 0.$$

Recalling $\phi(1) = \phi'(1) = 1$ and $\phi''(1) < \infty$, an order-two Taylor development of ϕ near $z = 1$ gives

$$\begin{aligned} \rho_{r+1} &= 1 + \phi'(1)(\rho_r - 1) + \frac{1}{2}\phi''(1)(\rho_r - 1)^2 \\ &= \rho_r + \frac{1}{2}\phi''(1)(\rho_r - 1)^2, \end{aligned}$$

leading to (with $a := \phi''(1)/2$) $\rho_r \sim 1 - \frac{1}{ar}$ as r is large.

The distribution of the extinction time $\tau_{m,0} := \inf(r > 0 : N_r(m) = 0)$ is therefore given by

$$\mathbf{P}(\tau_{m,0} \leq r) = \mathbf{P}(N_r(m) = 0) = [z^0] (\phi_r(z))^m = ([z^0] \phi_r(z))^m = \rho_r^m,$$

with Pareto(1) heavy tails

$$(12) \quad \mathbf{P}(\tau_{m,0} > r) \sim \frac{m}{ar} \text{ as } r \text{ is large.}$$

To summarize, $N_r(m)$ goes slowly to 0 with probability 1 (extinction is almost sure) but it takes a long time to do so.

Remark 4: All this is also true of course for the frame process $N_r(n)$ itself. As a critical branching process, the process $N_r(n)$ tends to 0 with probability 1 as well and it has a constant mean $\mathbf{E}(N_r(n)) = n$; it has a variance that goes to infinity linearly with r :

$$\sigma^2(N_r(n)) = rn\phi''(1).$$

It goes extinct with probability 1 but it takes a long but finite time $\tau_{n,0}$ to do so. We have

$$\mathbf{P}(\tau_{n,0} > r) = 1 - \phi_r(0)^n \sim n/(ra), \text{ for large } r,$$

with persistent heavy tails, non-geometric. In particular, $\mathbf{E}(\tau_{n,0}) = \infty$. The pgf of $N_r(n)$ conditioned on $N_r(n) > 0$ being

$$\frac{\phi_r(z)^n - \phi_r(0)^n}{1 - \phi_r(0)^n}, \text{ we get}$$

$$\mathbf{E}(N_r(n) \mid N_r(n) > 0) \sim ar, \text{ for large } r$$

with slow algebraic growth of order r (in r) and independent of n . Because $a = \phi''(1)/2 < \infty$, it holds as well that (Harris, 1964),

$$(13) \quad \mathbf{P}\left(\frac{N_r(n)}{ar} > x \mid N_r(n) > 0\right) \xrightarrow{r \rightarrow \infty} e^{-x}, \quad x > 0.$$

Therefore, consistently with the previous statements

$$\begin{cases} \mathbf{E}(N_r(n)) = 0 \cdot \mathbf{P}(\tau_{n,0} \leq r) + \mathbf{E}(N_r(n) \mid N_r(n) > 0) \mathbf{P}(\tau_{n,0} > r) \sim ar \frac{n}{ar} = n \\ \sigma^2(N_r(n)) = \sigma^2(N_r(n) \mid N_r(n) > 0) \mathbf{P}(\tau_{n,0} > r) \sim 2(ar)^2 \frac{n}{ar} = nr\phi''(1). \quad \square \end{cases}$$

Type 1 fixation event. The question of fixation of the descendance of the m founders is more tricky. We could define the fixation time of type 1 as $\tau_{m,n} := \inf(r > 0 : N_r(m) \geq n)$. But this definition is independent of what type 2 individuals do.

Rather we shall define the fixation event as follows. Suppose that the fixation time $\tau_{m,fix} = r$ if and only if generation r is the first time at which $N_r(n-m) = 0$ and $N_r(m) > 0$ (when type 2 goes extinct for the first time while some type 1 individuals still survive). Then, with $\rho_r = \mathbf{P}(N_r(1) = 0)$ obeying $\rho_0 = 0$ and $\rho_r \sim 1 - \frac{2}{r\phi''(1)} \rightarrow 1$ as $r \rightarrow \infty$,

$$(14) \quad \mathbf{P}(\tau_{m,fix} = r) = (1 - \rho_r^m)(\rho_r^{n-m} - \rho_{r-1}^{n-m}), \quad r \geq 1.$$

The probability that such a fixation event ever occurs therefore is

$$(15) \quad 0 < \mathbf{P}(\tau_{m,fix} < \infty) = \sum_{r \geq 1} \mathbf{P}(\tau_{m,fix} = r) < 1.$$

Note that

$$\mathbf{P}(\tau_{m,fix} \leq r, N_r(m) > 0) = (1 - \rho_r^m) \rho_r^{n-m}, \quad r \geq 1.$$

Clearly then, with probability $\mathbf{P}(\tau_{m,fix} < \infty)$, $\tau_{m,fix} < \tau_{m,0}$ and with complementary probability $\mathbf{P}(\tau_{m,fix} = \infty) = 1 - \mathbf{P}(\tau_{m,fix} < \infty)$, $\tau_{m,0} < \tau_{m,fix} = \infty$.

Type 1 population size at fixation. A natural question then is: given $\tau_{m,fix} < \infty$ (an event occurring with probability $\mathbf{P}(\tau_{m,fix} < \infty)$), what is the type 1 population size $N_{\tau_{m,fix}}(m)$ at the fixation time event? We have

$$\begin{aligned} \mathbf{E}(N_{\tau_{m,fix}}(m)) &= \sum_{r \geq 1} \mathbf{E}(N_r(m) \mid N_r(m) > 0) \mathbf{P}(\tau_{m,fix} = r) \\ &= m \sum_{r \geq 1} (\rho_r^{n-m} - \rho_{r-1}^{n-m}) = m, \end{aligned}$$

the sums telescoping with $\rho_0 = 0$ and $\rho_\infty = 1$. On average $N_{\tau_{m,fix}}(m)$ lies at m . Let us now consider the variance. We have

$$[\mathbf{E}(N_r(m) \mid N_r(m) > 0)]^2 = \frac{[\mathbf{E}(N_r(m))]^2}{(1 - \rho_r^m)^2} = \frac{m^2}{(1 - \rho_r^m)^2} \text{ and}$$

$$\mathbf{E}(N_r(m)^2 \mid N_r(m) > 0) = \frac{\mathbf{E}(N_r(m)^2)}{1 - \rho_r^m} = \frac{m^2 + rm\phi''(1)}{1 - \rho_r^m}.$$

Therefore,

$$\begin{aligned} \sigma^2(N_{\tau_{m,fix}}(m)) &= \sum_{r \geq 1} \sigma^2(N_r(m) \mid N_r(m) > 0) \mathbf{P}(\tau_{m,fix} = r) \\ &= \sum_{r \geq 1} \left[\frac{m^2 + rm\phi''(1)}{1 - \rho_r^m} - \frac{m^2}{(1 - \rho_r^m)^2} \right] \mathbf{P}(\tau_{m,fix} = r) \\ &= m\phi''(1) \sum_{r \geq 1} r (\rho_r^{n-m} - \rho_{r-1}^{n-m}) - m^2 \sum_{r \geq 1} \frac{\rho_r^m}{1 - \rho_r^m} (\rho_r^{n-m} - \rho_{r-1}^{n-m}) \\ &= m\phi''(1) \sum_{r \geq 1} r (\rho_r^{n-m} - \rho_{r-1}^{n-m}) + m^2 \mathbf{P}(\tau_{m,fix} = \infty). \end{aligned}$$

We obtained

$$(16) \quad \sigma^2(N_{\tau_{m,fix}}(m) \mid \tau_{m,fix} < \infty) = m\phi''(1) \sum_{r \geq 1} r (\rho_r^{n-m} - \rho_{r-1}^{n-m}).$$

Let us now show that the sum $\phi''(1) \sum_{r \geq 1} r (\rho_r^{n-m} - \rho_{r-1}^{n-m})$ is a divergent one. Recalling the asymptotic shape of ρ_r , the status of this sum is given by the status of the integral

$$\begin{aligned} I &= \int_1^\infty r\phi''(1) \frac{d}{dr} (\rho_r^{n-m}) dr \sim \int_1^\infty r\phi''(1) \frac{d}{dr} \left(\left(1 - \frac{2}{r\phi''(1)} \right)^{n-m} \right) dr \\ &= 2 \int_{\phi''(1)/2}^\infty s \frac{d}{ds} \left(\left(1 - \frac{1}{s} \right)^{n-m} \right) ds = 2(n-m) \int_{\phi''(1)/2}^\infty \frac{1}{s} \left(1 - \frac{1}{s} \right)^{n-m-1} ds \\ &= 2(n-m) \int_0^{2/\phi''(1)} \frac{1}{u} (1-u)^{n-m-1} du, \end{aligned}$$

which is indeed logarithmically diverging near 0. Thus $\sigma^2(N_{\tau_{m,fix}}(m)) = \infty$ and $N_{\tau_{m,fix}}(m)$ exhibits very large (infinite) fluctuations. The full pgf of $N_{\tau_{m,fix}}(m)$ clearly is

$$\begin{aligned} \mathbf{E}(z^{N_{\tau_{m,fix}}(m)}) &= \sum_{r \geq 1} \mathbf{E}(z^{N_r(m)} \mid N_r(m) > 0) \mathbf{P}(\tau_{m,fix} = r) \\ &= \sum_{r \geq 1} \frac{\phi_r(z)^m - \rho_r^m}{1 - \rho_r^m} \mathbf{P}(\tau_{m,fix} = r) \\ &= \sum_{r \geq 1} (\phi_r(z)^m - \rho_r^m) (\rho_r^{n-m} - \rho_{r-1}^{n-m}). \end{aligned}$$

Fixation probability and fixation time distribution. We can also estimate the fixation probability for $n \gg m$ large or for n, m large with $m/n = \alpha$. From (14), we have

$$(17) \quad \mathbf{P}(\tau_{m,fix} < \infty) = \sum_{r \geq 1} \mathbf{P}(\tau_{m,fix} = r) = \sum_{r \geq 1} (1 - \rho_r^m) (\rho_r^{n-m} - \rho_{r-1}^{n-m})$$

which can be approximated (observing $\rho_1 = \phi(0) = \pi_0$ and $\rho_r \sim 1 - \frac{2}{r\phi''(1)}$) by

$$\begin{aligned} I &= \int_1^\infty dr (1 - \rho_r^m) \frac{d}{dr} (\rho_r^{n-m}) = (n-m) \int_1^\infty d\rho_r (1 - \rho_r^m) \rho_r^{n-m-1} \\ &= (n-m) \int_{\pi_0}^1 d\rho (1 - \rho^m) \rho^{n-m-1} = \frac{m}{n} (1 - \pi_0^n) - \pi_0^{n-m}. \end{aligned}$$

of order m/n when n goes large at fixed m or when both m, n go large at fixed $m/n = \alpha$. This estimate makes sense because in these two cases, the factor ρ_r^{n-m-1} goes to 0 for small values of r so that only the large r behavior of ρ contributes where the asymptotic estimate of ρ_r is valid. With $a := \phi''(1)/2$, we also have

$$\begin{aligned} \mathbf{E}(\tau_{m,fix} \mid \tau_{m,fix} < \infty) &= I^{-1} \sum_{r \geq 1} r \mathbf{P}(\tau_{m,fix} = r, \tau_{m,fix} < \infty) \\ (18) \quad &\sim \frac{(n-m)}{I} \int_1^\infty d\rho_r r (1 - \rho_r^m) \rho_r^{n-m-1} \sim \frac{n-m}{aI} \int_{\pi_0}^1 d\rho \frac{(1 - \rho^m) \rho^{n-m-1}}{1 - \rho} \\ &= \frac{n-m}{aI} \sum_{k=1}^m \frac{1}{n-m-k-1} (1 - \pi_0^{n-m-k-1}) < \infty. \end{aligned}$$

Note that, proceeding similarly

$$(19) \quad \mathbf{E}(\tau_{m,fix}^2 \mid \tau_{m,fix} < \infty) \sim \frac{n-m}{aI} \int_{\pi_0}^1 d\rho \frac{1 - \rho^m}{(1 - \rho)^2} \rho^{n-m-1},$$

which is a diverging integral near $\rho = 1$. The variance of $\tau_{m,fix} \mid \tau_{m,fix} < \infty$ is infinite.

The formula (18) giving the mean fixation time allows to obtain asymptotic estimates of this quantity, given fixation occurs. If n goes large at fixed m , we indeed find

$$(20) \quad \mathbf{E}(\tau_{m,fix} \mid \tau_{m,fix} < \infty) \sim \frac{n^2}{ma} \frac{m}{n} = 2n/\phi''(1).$$

If both m, n get large at fixed $m/n = \alpha \in (0, 1)$, we obtain

$$(21) \quad \begin{cases} \mathbf{E}(\tau_{m,fix} \mid \tau_{m,fix} < \infty) \sim \frac{n-m}{aI} \int_0^{m-1} dk \frac{1 - \pi_0^{n(1-\rho)-k}}{n(1-\alpha)-k} \\ \sim -\frac{2n}{\phi''(1)} \frac{1-\alpha}{\alpha} \log(1-\alpha). \end{cases}$$

A symmetric definition of $\tau_{m,ext}$ and $\tau_{m,fix}$ for type 1. We could have defined alternatively (and more symmetrically) the extinction time of type 1 as

$$\tau_{m,ext} := \inf(r > 0 : N_r(m) = 0, N_r(n-m) > 0),$$

so as the fixation time of type 2 individuals. And this definition is now dependent of what type 2 individuals do. We shall let $\tau_m = \tau_{m,ext} \wedge \tau_{m,fix}$, the global absorption time. To compute the probability that $\tau_{m,ext} < \infty$ from this new definition of

$\tau_{m,ext}$, it suffices to substitute $n - m$ to m in the expression of $\mathbf{P}(\tau_{m,fix} < \infty)$ for type 1 individuals as computed in (17).

And clearly, $\tau_{m,ext} \vee \tau_{m,fix} = \infty$, $\tau_{m,ext} < \infty \Rightarrow \tau_{m,ext} < \tau_{m,fix} = \infty$ and $\tau_{m,fix} < \infty \Rightarrow \tau_{m,fix} < \tau_{m,ext} = \infty$.

We have $\mathbf{P}(\tau_{m,fix} < \infty) + \mathbf{P}(\tau_{m,ext} < \infty) < 1$, and the missing mass is the probability that we have neither a fixation nor an extinction event of type 1, which happens whenever both type 1 and 2 populations die out simultaneously. This (rare) event occurs with probability

$$\theta := \sum_{r \geq 1} (\rho_r^m - \rho_{r-1}^m) (\rho_r^{n-m} - \rho_{r-1}^{n-m}),$$

and it is a case when both $\tau_{m,ext}$ and $\tau_{m,fix} = \infty$ (neither fixation nor extinction are achieved and $\theta = \mathbf{P}(\tau_{m,ext} = \tau_{m,fix} = \infty)$). Yet, given this last event has not occurred, we now know how to compute the probability that a fixation event of type 1 precedes an extinction event and conversely, normalizing the unconditional fixation/extinction probabilities by $1 - \theta$.

Define next

$$\mathbf{P}(k'_1, k' \mid k_1, k) := \mathbf{P}(N_{r+1}(m) = k'_1, N_{r+1}(n) = k' \mid N_r(m) = k_1, N_r(n) = k) =$$

$$\mathbf{P}(\xi_1 + \dots + \xi_{k_1} = k'_1) \mathbf{P}(\xi_{k_1+1} + \dots + \xi_k = k' - k'_1) = \left[z^{k'_1} \right] \phi(z)^{k_1} \left[z^{k' - k'_1} \right] \phi(z)^{k - k_1}.$$

Note that $k_1 = k \Rightarrow \mathbf{P}(k'_1, k' \mid k_1, k) = \left[z^{k'} \right] \phi(z)^k \cdot \delta_{k'}(k'_1)$ (type 1 is fixed), whereas $k_1 = 0 \Rightarrow \mathbf{P}(k'_1, k' \mid k_1, k) = \left[z^{k'} \right] \phi(z)^k \cdot \delta_0(k'_1)$ (type 2 is fixed).

We have

$$\begin{aligned} \mathbf{P}(N_{r+1}(m) = k'_1 \mid N_r(m) = k_1, N_r(n) = k) &= \sum_{k' \geq k'_1} \mathbf{P}(k'_1, k' \mid k_1, k) \\ &=: \mathbf{P}(k'_1 \mid k_1, k) = \left(\left[z^{k'_1} \right] \phi(z)^{k_1} \right) \phi(1)^{k - k_1} = \left[z^{k'_1} \right] \phi(z)^{k_1}, \end{aligned}$$

independent of k . Thus $\mathbf{E}(z^{N_{r+1}(m)} \mid k_1, k) = \phi(z)^{k_1}$, independent of k . In particular,

$$\begin{aligned} \mathbf{E}(N_{r+1}(m) \mid k_1, k) &= \sum_{k'_1} k'_1 \left[z^{k'_1} \right] \phi(z)^{k_1} = \phi(1) k_1 \phi'(1) = k_1 = k \frac{k_1}{k} \\ \sigma^2(N_{r+1}(m) \mid k_1, k) &= k_1 \phi''(1) = k \frac{k_1}{k} \phi''(1), \end{aligned}$$

and $N_r(m)$ is a martingale. Similarly, using the convolution formula,

$$\begin{aligned} \mathbf{P}(N_{r+1}(n) = k' \mid N_r(m) = k_1, N_r(n) = k) &= \sum_{k'_1 \leq k'} \mathbf{P}(k'_1, k' \mid k_1, k) \\ &=: \mathbf{P}(k' \mid k_1, k) = \sum_{0 \leq k'_1 \leq k'} \left[z^{k'_1} \right] \phi(z)^{k_1} \left[z^{k' - k'_1} \right] \phi(z)^{k - k_1} = \left[z^{k'} \right] \phi(z)^k, \end{aligned}$$

independent of k_1 . This illustrates that both $N_r(m)$ and $N_r(n)$, although not independent, are both Markov processes.

We also have consistently

$$\begin{aligned}
 \mathbf{E} \left(z_1^{N_{r+1}(m)} z^{N_{r+1}(n)} \mid k_1, k \right) &= \sum_{k'_1 \geq 0} z_1^{k'_1} \left[z_1^{k'_1} \right] \phi(z_1)^{k_1} \sum_{k' \geq k'_1} z^{k'} \left[z^{k'-k'_1} \right] \phi(z)^{k-k_1} \\
 (22) \quad &= \sum_{k'_1 \geq 0} (zz_1)^{k'_1} \left[z_1^{k'_1} \right] \phi(z_1)^{k_1} \phi(z)^{k-k_1} = \phi(zz_1)^{k_1} \phi(z)^{k-k_1},
 \end{aligned}$$

translating the fact that $N_r(m)$, $N_r(n)$ are jointly Markov as well (with Markov marginals).

Introducing bias (deviation to neutrality): We now wish to introduce bias translating some interaction between type 1 and type 2 populations and see what changes as compared to the neutral case. For some bias function $p : [0, 1] \rightarrow [0, 1]$, different from the identity, such as the ones introduced in Section 3, let

$$\begin{aligned}
 \mathbf{P}(N_{r+1}(m) = k'_1, N_{r+1}(n) = k' \mid N_r(m) = k_1, N_r(n) = k) \\
 (23) \quad &= \left[z^{k'_1} \right] \phi(z)^{kp(k_1/k)} \left[z^{k'-k'_1} \right] \phi(z)^{k(1-p(k_1/k))},
 \end{aligned}$$

be the transition probability matrix of the joint process $N_r(m)$, $N_r(n)$, now with bias p . Then

$$\begin{aligned}
 \mathbf{P}(N_{r+1}(m) = k'_1 \mid N_r(m) = k_1, N_r(n) = k) &= \sum_{k' \geq k'_1} \mathbf{P}(k'_1, k' \mid k_1, k) \\
 &=: \mathbf{P}(k'_1 \mid k_1, k) = \left(\left[z^{k'_1} \right] \phi(z)^{kp(k_1/k)} \right) \phi(1)^{k(1-p(k_1/k))} = \left[z^{k'_1} \right] \phi(z)^{kp(k_1/k)},
 \end{aligned}$$

Thus $\mathbf{E}(z_1^{N_{r+1}(m)} \mid k_1, k) = \phi(z_1)^{kp(k_1/k)}$, now dependent on both k_1, k . In particular,

$$(24) \quad \begin{cases} \mathbf{E}(N_{r+1}(m) \mid k_1, k) = kp(k_1/k) \\ \sigma^2(N_{r+1}(m) \mid k_1, k) = kp(k_1/k) \phi''(1). \end{cases}$$

The marginal process $N_r(m)$ is no longer Markov, nor is it a martingale anymore as a result of introducing bias. We also have, consistently,

$$\begin{aligned}
 \mathbf{E} \left(z_1^{N_{r+1}(m)} z^{N_{r+1}(n)} \mid k_1, k \right) &= \\
 &= \sum_{k'_1 \geq 0} z_1^{k'_1} \left[z_1^{k'_1} \right] \phi(z_1)^{kp(k_1/k)} \sum_{k' \geq k'_1} z^{k'} \left[z^{k'-k'_1} \right] \phi(z)^{k(1-p(k_1/k))} = \\
 &\sum_{k'_1 \geq 0} (zz_1)^{k'_1} \left[z_1^{k'_1} \right] \phi(z_1)^{kp(k_1/k)} \phi(z)^{k(1-p(k_1/k))} = \phi(zz_1)^{kp(k_1/k)} \phi(z)^{k(1-p(k_1/k))},
 \end{aligned}$$

translating that $N_r(m)$, $N_r(n)$ are still jointly Markov though. Note $\mathbf{E}(z^{N_{r+1}(n)} \mid k_1, k) = \phi(z)^k$, independent of k_1 : the frame process $N_r(n)$ still is a Markov critical Galton-Watson process. And, with $N_r(n-m) = N_r(n) - N_r(m)$, given $N_r(n-m) = k_2 = k - k_1$ and $N_r(n) = k$, $\mathbf{E}(z_2^{N_{r+1}(n-m)} \mid k_2, k) = \phi(z_2)^{k(1-p(1-k_2/k))}$, dependent on both k_2, k . Type 2 $N_r(n-m)$ is non-Markov either.

Effective population size: when n is a fixed population size, $c_n = O(1/n)$ (as in the Pólya model for example) is the probability that two distinct randomly chosen individuals out of n have the same direct ancestor. If the population profile is

$N_r(n)$ and so variable, c_n should be replaced by $\bar{c}_n = R^{-1} \sum_{r=1}^R 1/N_r(n)$, the empirical average of the $c_{N_r(n)}$ s over some (maybe long) time period R . The effective population size $n_e = 1/c_n$ should accordingly be replaced by

$$n_e = 1/\bar{c}_n = \frac{R}{\sum_{r=1}^R 1/N_r(n)},$$

the harmonic mean of the $N_r(n)$ s. The effective population size used when there are changes in population sizes is thus the harmonic mean (rather than the arithmetic one). This reflects that if a population has recovered from a bottleneck (a dramatic reduction in population size), a definition of the effective population size after a bottleneck should enhance instants where $N_r(n)$ is small. Because genetic drift acts more quickly to reduce genetic variation in small populations (see Remark 2), genetic diversity in a population will indeed be substantially reduced when the population size shrinks in a bottleneck event (the founder effect).

Our construction with variable population size is conditionally given $N_0(n) = n$, where $N_r(n)$ is a critical Galton-Watson sequence of random variables giving the random size of the global population in each generation r . Averaging over the $N_r(n)$ s, while taking into account that $N_r(n)$ goes extinct with probability 1 at time $\tau_{n,0} = \inf(r \geq 1 : N_r(n) = 0)$, we can define the effective population size n_e in our context by

$$(25) \quad n_e = \mathbf{E} \frac{\tau_{n,0}}{\sum_{r=0}^{\tau_{n,0}-1} 1/N_r(n)} < \mathbf{E} \frac{\sum_{r=0}^{\tau_{n,0}-1} N_r(n)}{\tau_{n,0}} = an \mathbf{P}(\tau_{n,0} < \infty) = an.$$

We used here that the harmonic mean is dominated by the arithmetic one.

Survival probability to a bottleneck effect. A bottleneck effect occurs when the total population size shrinks to a small value before recovering. We wish to compute the survival probability of type 1 individuals in such a situation. More precisely, let k_1, k_2 obeying $k_1 \ll k_2$ be the (bottom and top) total population sizes at $r_1 < r_2$, respectively. We may assume k_2 of order n , the mean value of the total population size at r_2 , whereas k_1 is assumed comparatively small because of shrinkage at r_1 . We therefore wish to compute

$$\mathbf{P}(N_{r_2}(m) > 0 \mid N_{r_1}(n) = k_1, N_{r_2}(n) = k_2).$$

We have

$$\mathbf{E} \left(z_1^{N_{r_1}(m)} z_2^{N_{r_2}(m)} z_3^{N_{r_1}(n-m)} z_4^{N_{r_2}(n-m)} \right) = \phi_{r_1} \left(z_1 \phi_{r_2-r_1}(z_2) \right)^m \phi_{r_1} \left(z_3 \phi_{r_2-r_1}(z_4) \right)^{n-m},$$

therefore (substituting $z_2 z_4$ to z_2 and taking $z_1 = z_3$)

$$(26) \quad \begin{aligned} \Phi_{r_1, r_2}(z_2, z_3, z_4) &:= \mathbf{E} \left(z_2^{N_{r_2}(m)} z_3^{N_{r_1}(n)} z_4^{N_{r_2}(n)} \right) \\ &= \phi_{r_1} \left(z_3 \phi_{r_2-r_1}(z_2 z_4) \right)^m \phi_{r_1} \left(z_3 \phi_{r_2-r_1}(z_4) \right)^{n-m}. \end{aligned}$$

We thus get

$$\begin{aligned} &\mathbf{P}(N_{r_2}(m) > 0 \mid N_{r_1}(n) = k_1, N_{r_2}(n) = k_2) \\ &= 1 - \frac{\left[z_2^0 z_3^{k_1} z_4^{k_2} \right] \Phi_{r_1, r_2}(z_2, z_3, z_4)}{\left[z_3^{k_1} z_4^{k_2} \right] \Phi_{r_1, r_2}(1, z_3, z_4)} = 1 - \frac{\left[z_3^{k_1} z_4^{k_2} \right] \Phi_{r_1, r_2}(0, z_3, z_4)}{\left[z_3^{k_1} z_4^{k_2} \right] \Phi_{r_1, r_2}(1, z_3, z_4)}, \end{aligned}$$

where

$$\begin{aligned}\Phi_{r_1, r_2}(1, z_3, z_4) &= \phi_{r_1}(z_3 \phi_{r_2-r_1}(z_4))^n \\ \Phi_{r_1, r_2}(0, z_3, z_4) &= \phi_{r_1}(z_3 \phi_{r_2-r_1}(0))^m \phi_{r_1}(z_3 \phi_{r_2-r_1}(z_4))^{n-m}.\end{aligned}$$

We have

$$\begin{aligned}\begin{bmatrix} z_3^{k_1} \end{bmatrix} \Phi_{r_1, r_2}(1, z_3, z_4) &= \phi_{r_2-r_1}(z_4)^{k_1} \cdot \begin{bmatrix} z_3^{k_1} \end{bmatrix} \phi_{r_1}(z_3)^n \\ \begin{bmatrix} z_3^{k_1} z_4^{k_2} \end{bmatrix} \Phi_{r_1, r_2}(1, z_3, z_4) &= \begin{bmatrix} z_3^{k_1} \end{bmatrix} \phi_{r_1}(z_3)^n \cdot \begin{bmatrix} z_4^{k_2} \end{bmatrix} \phi_{r_2-r_1}(z_4)^{k_1} \\ \begin{bmatrix} z_3^{k_1} \end{bmatrix} \Phi_{r_1, r_2}(0, z_3, z_4) &= \sum_{k=0}^{k_1} \begin{bmatrix} z_3^k \end{bmatrix} \phi_{r_1}(z_3 \phi_{r_2-r_1}(z_4))^{n-m} \cdot \begin{bmatrix} z_3^{k_1-k} \end{bmatrix} \phi_{r_1}(z_3 \phi_{r_2-r_1}(0))^m \\ &= \sum_{k=0}^{k_1} \phi_{r_2-r_1}(z_4)^k \begin{bmatrix} z_3^k \end{bmatrix} \phi_{r_1}(z_3)^{n-m} \phi_{r_2-r_1}(0)^{k_1-k} \begin{bmatrix} z_3^{k_1-k} \end{bmatrix} \phi_{r_1}(z_3)^m \\ \begin{bmatrix} z_3^{k_1} z_4^{k_2} \end{bmatrix} \Phi_{r_1, r_2}(0, z_3, z_4) &= \\ \sum_{k=0}^{k_1} \begin{bmatrix} z_4^{k_2} \end{bmatrix} \phi_{r_2-r_1}(z_4)^k \cdot \begin{bmatrix} z_3^k \end{bmatrix} \phi_{r_1}(z_3)^{n-m} \phi_{r_2-r_1}(0)^{k_1-k} \cdot \begin{bmatrix} z_3^{k_1-k} \end{bmatrix} \phi_{r_1}(z_3)^m.\end{aligned}$$

We thus only need to compute $\begin{bmatrix} z^k \end{bmatrix} \phi_r(z)^n$ for various values of k, n ¹. With $f(z) = (\phi_r(0) + z)^n$, $g(z) = \phi_r(z) - \phi_r(0)$, we have

$$\phi_r(z)^n = f(g(z)).$$

By Faà di Bruno formula (see e.g. Comtet, 1970, Tome 1, p. 148), with $B_{k,l}(g_1, g_2, \dots)$ the Bell polynomials in the Taylor coefficients g_k of $g(z)$: $g_k = k! \begin{bmatrix} z^k \end{bmatrix} g(z) = k! \mathbf{P}(N_r(1) = k)$,

$$\begin{bmatrix} z^k \end{bmatrix} \phi_r(z)^n =: a_{k,n}(r) = \frac{1}{k!} \sum_{l=1}^k (n)_l \phi_r(0)^{n-l} B_{k,l}(g_1, g_2, \dots).$$

So we end up with a combinatorial closed form formula for the survival probability to a bottleneck effect as²:

$$(27) \quad 1 - \frac{\mathbf{P}(N_{r_2}(m) > 0 \mid N_{r_1}(n) = k_1, N_{r_2}(n) = k_2) = \sum_{k=0}^{k_1} a_{k_2,k}(r_2-r_1) a_{k,n-m}(r_1) a_{k_1-k,m}(r_1) a_{0,1}(r_2-r_1)^{k_1-k}}{a_{k_1,n}(r_1) a_{k_2,k_1}(r_2-r_1)}.$$

Example 10 With $p_0 + q_0 = p + q = 1$, suppose the critical homographic model $\phi(z) = q_0 + p_0 \frac{qz}{1-pz}$, a mixture of a Bernoulli(p_0) random variable with a geometric(p) one. We have $\phi'(1) = 1$ if $p_0 = q$ and then $\phi''(1) = 2p/q =: 2a$. Thus

¹Following (Gardy, 1995), useful asymptotic estimates of such large powers quantities under different regimes for k, n are available, namely: k fixed, $n \rightarrow \infty$, $k = o(n)$ and $n \rightarrow \infty$ or $k = O(n)$ and $n \rightarrow \infty$.

²Observing: $\phi_r(0) := \begin{bmatrix} z^0 \end{bmatrix} \phi_r(z) = a_{0,1}(r)$.

$$\begin{aligned}\phi(z) &= 1 - \left((1-z)^{-1} + a\right)^{-1} \text{ and} \\ \phi_r(z) &= 1 - \left((1-z)^{-1} + ra\right)^{-1} \text{ or } \phi_r(z) = \frac{ra + z(1-ra)}{1 + ra - raz}.\end{aligned}$$

Here, $[z^0] \phi_r(z)^n = \left(\frac{ra}{1+ra}\right)^n$ and if $k \geq 1$,

$$(28) \quad \begin{cases} [z^k] \phi_r(z)^n =: a_{k,n}(r) \\ = \sum_{l=1}^n \binom{n}{l} \binom{k+l-1}{k} \left(\frac{ra}{1+ra}\right)^k \left(\frac{ra-1}{ra}\right)^{n-l} (ra(1+ra))^{-l} \\ = \left(\frac{ra}{ra+1}\right)^k \left(\frac{ra-1}{ra}\right)^n \sum_{l=1}^n \binom{n}{l} \binom{k+l-1}{k} \left((ra)^2 - 1\right)^{-l}, \end{cases}$$

which can be expressed in terms of an hypergeometric function ${}_2F_1$.

This homographic (or linear-fractional) model is of particular interest in our context for two reasons:

1/ One is because $\phi_r(z)$, as the r -th composition of $\phi(z)$ with itself, is explicitly computable, as is well-known from the theory of branching processes (the invariance under iterated composition property; Harris, 1963, p. 9). $\phi(z)$ is the pgf of an infinitely divisible random variable ξ if and only if $p \geq \sqrt{2} - 1$ or $\sigma^2(\xi) = \phi''(1) \geq \sqrt{2}$ (Steutel and van Harn, Example 11.15, 2003).

2/ Because all critical branching processes generated by some random variable ξ are in the domain of attraction of the critical homographic one. Indeed, with $\phi(z)$ the pgf of ξ , the pgf $\phi_r(z)$ of $N_r(1)$ obeys

$$\phi_{r+1}(z) = \phi(\phi_r(z)), \phi_0(z) = z.$$

When r gets large, due to almost sure extinction, $\phi_r(z)$ approaches 1 (the pgf of $N_r(1) = 0$). Recalling $\phi(1) = \phi'(1) = 1$ and $\phi''(1) < \infty$, an order-two Taylor development of ϕ near $z = 1$ therefore gives

$$\begin{aligned}\phi_{r+1}(z) &= 1 + \phi'(1)(\phi_r(z) - 1) + \frac{1}{2}\phi''(1)(\phi_r(z) - 1)^2 \\ &= \phi_r(z) + \frac{1}{2}\phi''(1)(\phi_r(z) - 1)^2,\end{aligned}$$

leading (recalling $a := \phi''(1)/2$) to

$$\phi_r(z) \sim 1 - 1/\left((1-z)^{-1} + ra\right), \text{ as } r \text{ is large. } \square$$

Acknowledgments:

T. Huillet acknowledges partial support from the “Chaire *Modélisation mathématique et biodiversité*”. N. Grosjean and T. Huillet also acknowledge support from the labex MME-DII Center of Excellence (*Modèles mathématiques et économiques de la dynamique, de l'incertitude et des interactions*, ANR-11-LABX-0023-01 project).

REFERENCES

- [1] Cannings, C. The latent roots of certain Markov chains arising in genetics: a new approach. I. Haploid models. *Advances in Appl. Probability* 6, 260–290 (1974).
- [2] Comtet, L. *Analyse combinatoire*. Tomes 1 et 2. Presses Universitaires de France, Paris, 1970.
- [3] Crow, J.; F.; Kimura, M. *An introduction to population genetics theory*. Harper & Row, Publishers, New York-London, 1970.

- [4] Ewens, W. J. *Mathematical population genetics. I. Theoretical introduction*. Second edition. Interdisciplinary Applied Mathematics, 27. Springer-Verlag, New York, 2004.
- [5] Gardy, D. Some results on the asymptotic behaviour of coefficients of large powers of functions. *Discrete Mathematics*. 139(13), 189-217, (1995).
- [6] Gillespie, J. H. *The Causes of Molecular Evolution*. New York and Oxford: Oxford University Press, 1991.
- [7] Gladstien, K. The characteristic values and vectors for a class of stochastic matrices arising in genetics. *SIAM J. Appl. Math.* 34, no. 4, 630-642, (1978).
- [8] Harris, T. E. *The theory of branching processes*. Die Grundlehren der Mathematischen Wissenschaften, Bd. 119 Springer-Verlag, Berlin; Prentice-Hall, Inc., Englewood Cliffs, N.J., 1963.
- [9] Huillet, T.; Möhle, M. Asymptotics of symmetric compound Poisson population models. *Combinatorics, Probability and Computing*, 24, 216-253, (2015).
- [10] Karlin, S.; McGregor, J. Direct product branching processes and related Markov chains. *Proc. Nat. Acad. Sci. U.S.A.*, 51, 598-602, (1964).
- [11] Kingman, J. F. C. The coalescent. *Stochastic Process. Appl.*, 13, 235-248, (1982).
- [12] Liggett, T. M. *Interacting particle systems*. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 276. Springer-Verlag, New York, 1985.
- [13] Maruyama, T. *Stochastic problems in population genetics*. Lecture Notes in Biomathematics, 17. Springer-Verlag, Berlin-New York, 1977.
- [14] Möhle, M. Forward and backward processes in bisexual models with fixed population sizes. *J. Appl. Probab.* 31, no. 2, 309-332, (1994).
- [15] Möhle, M. The concept of duality and applications to Markov processes arising in neutral population genetics models. *Bernoulli* 5, no. 5, 761-777, (1999).
- [16] Nagylaki, T. Anecdotal, historical and critical commentaries on Genetics. Gustave Malécot and the transition from classical to modern population genetics. Edited by James F. Crow and William F. Dove. *Genetics*, 122, 253-268, (1989).
- [17] Steutel, F. W.; van Harn, K. *Infinite Divisibility of Probability Distributions on the Real Line*. Chapman & Hall/CRC Pure and Applied Mathematics, 2003.
- [18] Tavaré, S. Ancestral inference in population genetics. Lectures on probability theory and statistics, Saint-Flour 2001, *Lecture Notes in Math.*, 1837, 1-188, Springer, (2004).
- [19] Weissing, F. J., van Boven M. Selection and segregation distortion in a sex-differentiated population. *Theor. Popul. Biol.*, 60(4), 327-41, (2001).

LABORATOIRE DE PHYSIQUE THÉORIQUE ET MODÉLISATION CNRS-UMR 8089 ET UNIVERSITÉ DE CERGY-PONTOISE, 2 AVENUE ADOLPHE CHAUVIN, 95302, CERGY-PONTOISE, FRANCE, E-MAIL(S): NICOLAS.GROSJEAN@U-CERGY.FR AND THIERRY.HUILLET@U-CERGY.FR